Bayes Factor Functions

Saptati Datta, Rachael Shudde, Valen E. Johnson

Texas A&M University, College Station, TX

October 25, 2023

Outline

- Practical issues with Bayes factors
- Bayes factors based on test statistics (BFBOTS)
- Non-local Alternative Prior Densities (NAPs)
- Bayes factor functions (BFFs)
- Examples

Practical issues with Bayes factors

- Defining null and alternative models is difficult in high-dimensional settings.

Site	Results for the following blood groups		
	O	A	B or AB
Pylorus and antrum	104	140	52
Body and fundus	116	117	52
Cardia	28	39	11
Extensive	28	12	8

White and Eisenberg's classification of cancer patients.

Practical issues

Bayes factors based on test statistics (BFBOTS)[Proposed by V.E. Johnson(2005)]

- Suppose X is a standard test statistic (i.e., z, t, χ^2, F).
  - Under H_0, distribution, distribution of test statistic is known. No prior densities are needed.
  - Under H_1, distribution of X depends on scalar non-centrality parameter. Only prior on scalar needed.
  - Avoids high-dimensional integration.
  - Avoids high-dimensional prior specification.

Practical issues: BFBOTS

Practical issues: BFBOTS

- BFBOTS do not reflect effect sizes

Practical issues: BFBOTS

- BFBOTS do not reflect effect sizes
- Bayes Factors expressed as functions of effect sizes already proposed(V.E. Johnson, S. Pramanik, R.Shudde, PNAS 2023).

Practical issues: BFBOTS

- BFBOTS do not reflect effect sizes
- Bayes Factors expressed as functions of effect sizes already proposed(V.E. Johnson, S. Pramanik, R.Shudde, PNAS 2023).
- **Aim of this project:** To account for the variability of the effect sizes through another hyper-parameter

Non-local priors

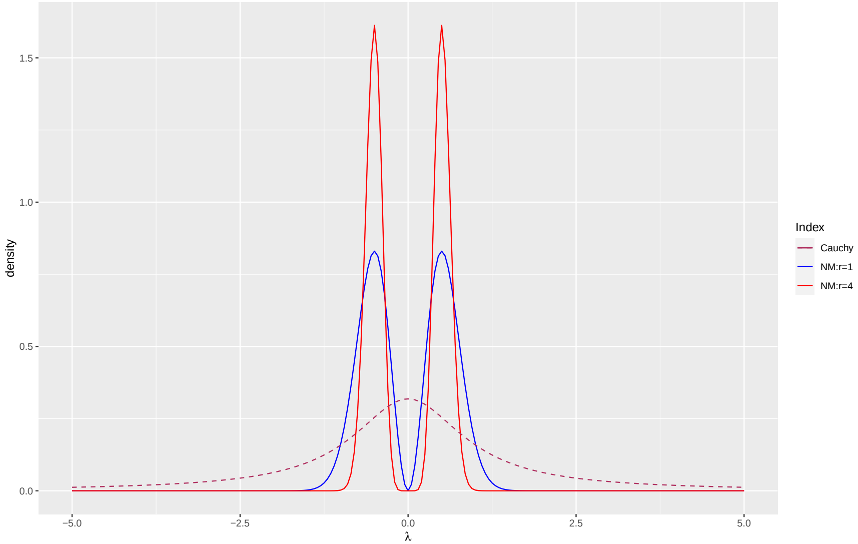
- Define alternative priors so that they assign negligible mass to parameters consistent with the null hypothesis
- For normal mean, the prior on μ has a prior density defined as follows

$$\pi_{NM}(\mu \,|\, r, \tau^2) = \frac{(\mu^2)^r}{(2\tau^2)^{r+\frac{1}{2}}\Gamma\left(r+\frac{1}{2}\right)} \exp\left(-\frac{\mu^2}{2\tau^2}\right), \quad \mu \in \mathrm{R}, \quad \tau, r > 0$$

.
is a normal moment (NM) prior density,

- $\pi_{NM}(0 \,|\, \tau^2) = 0$
- Modes are at $\pm\sqrt{2r}\tau$

Example: JZS and NM priors for normal mean



Z test

Theorem. *Assume the distributions of a random variable z under the null and alternative hypotheses are described by*

$$\begin{aligned} H_0 : z &\sim N(0,1), \\ H_1 : z \,|\, \lambda &\sim N(\lambda,1), \qquad \lambda \,|\, \tau^2 \sim NM(r,\tau^2). \end{aligned}$$

Then the Bayes factor in favor of the alternative hypothesis is

$$BF_{10}(z|\tau^2) = \frac{1}{(1+\tau^2)^{r+\frac{1}{2}}}\,{}_1F_1\left(r+\frac{1}{2}, \frac{1}{2}; \frac{\tau^2 z^2}{2(1+\tau^2)}\right) \tag{1}$$

T test

Theorem *Assume the distributions of a random variable* t_ν *under the null and alternative hypotheses are described by*

$$
\begin{aligned}
H_0 : t &\sim T_\nu(0), \\
H_1 : t \,|\, \lambda &\sim T_\nu(\lambda), \qquad \lambda \,|\, \tau^2 \sim NM(r, \tau^2).
\end{aligned}
$$

$$
\begin{aligned}
& BF_{10}(t \mid \tau^2, r) \\
&= \frac{1}{(1+\tau^2)^{r+\frac{1}{2}}} {}_2F_1\left(\frac{\nu+1}{2}, r+\frac{1}{2}, \frac{1}{2}, \frac{\tau^2 t^2}{(t^2+\nu)(\tau^2+1)}\right) \\
&+ \frac{t\tau}{\sqrt{t^2+\nu}(\tau^2+1)^{r+1}} \frac{\Gamma(\frac{\nu}{2}+1)}{\Gamma\left(\frac{\nu+1}{2}\right)} \frac{\Gamma(r+1)}{\Gamma(r+\frac{1}{2})} {}_2F_1\left(\frac{\nu}{2}+1, r+1, \frac{3}{2}, \frac{t^2\tau^2}{(t^2+\nu)(1+\tau^2)}\right) \qquad (2)
\end{aligned}
$$

χ^2 test

Theorem *Assume the distributions of a random variable h under the null and alternative hypotheses are described by*

$$\begin{aligned} H_0 : h &\sim \chi_k^2(0), \\ H_1 : h \,|\, \lambda &\sim \chi_k^2(\lambda), \qquad \lambda \,|\, \tau^2 \sim G\left(\frac{k}{2} + r, \frac{1}{2\tau^2}\right). \end{aligned}$$

Then the Bayes factor in favor of the alternative hypothesis is

$$BF_{10}(h \mid \tau^2) = \frac{1}{(1+\tau^2)^{k/2+r}} \, {}_1F_1\left(\frac{k}{2} + r, \frac{k}{2}; \frac{\tau^2 h}{2(1+\tau^2)}\right) \tag{3}$$

F test

Theorem *Assume the distributions of a random variable f under the null and alternative hypotheses are described by*

$$\begin{aligned} H_0 : f &\sim F_{k,m}(0), \\ H_1 : f \,|\, \lambda &\sim F_{k,m}(\lambda), \quad \lambda \,|\, \tau^2 \sim G\left(\frac{k}{2} + r, \frac{1}{2\tau^2}\right). \end{aligned}$$

Then the Bayes factor in favor of the alternative hypothesis is

$$BF_{10}(f \mid \tau^2) = \frac{1}{(1+\tau^2)^{\frac{k}{2}+r}} {}_2F_1(k/2+r, \frac{k+m}{2}, k/2; \frac{kf\tau^2}{(1+\tau^2)(m+kf)}) \quad (4)$$

where $v = m(\tau^2+1)$.

Rates of Convergence

- **Z-test:** Suppose that the following hold:

  (i) $z \sim N(\gamma\sqrt{n}, 1)$ when H_1 is true,

  (ii) $\tau^2 = \beta n$ for $\beta > 0$.

  Then $BF_{01}(z \,|\, \tau^2, r) = O_p(\exp(-cn))$ for some $c > 0$ when H_1 applies, and $BF_{10}(z \,|\, \tau^2, r) = O_p(n^{r+\frac{1}{2}})$ when H_0 is true.

Rates of Convergence

- **Z-test:** Suppose that the following hold:

  (i) $z \sim N(\gamma\sqrt{n}, 1)$ when H_1 is true,

  (ii) $\tau^2 = \beta n$ for $\beta > 0$.

  Then $BF_{01}(z \,|\, \tau^2, r) = O_p(\exp(-cn))$ for some $c > 0$ when H_1 applies, and $BF_{10}(z \,|\, \tau^2, r) = O_p(n^{r+\frac{1}{2}})$ when H_0 is true.

- χ^2 **test:** Suppose the following hold:

  (i) $h \sim \chi_k^2(\gamma n)$ for some $\gamma > 0$ when the alternative hypothesis is true, and

  (ii) $\tau^2 = \beta n$ for some $\beta > 0$.

  Then $BF_{01}(h \mid \tau^2, r) = O_p[\exp(-cn)]$ for some $c > 0$ when the alternative hypothesis is true and $BF_{10}(h \mid \tau^2, r) = O_p(n^{-r-\frac{k}{2}})$ when the null hypothesis is true.

Choice of τ^2: Reflects effect size

Choice of τ^2: Reflects effect size

- Assume r is known for a single study.

Choice of τ^2: Reflects effect size

- Assume r is known for a single study.
- Express the mode of the non-centrality parameter as a function of the standardized effect-size, say $\psi(\omega, r)$.

$$\psi(\omega, r) = \arg\max_{\lambda} \pi(\lambda \mid \tau^2_{\omega,r}), \tag{5}$$

Choice of τ^2: Reflects effect size

- Assume r is known for a single study.
- Express the mode of the non-centrality parameter as a function of the standardized effect-size, say $\psi(\omega, r)$.

$$\psi(\omega, r) = \arg\max_{\lambda} \pi(\lambda \mid \tau^2_{\omega,r}), \quad (5)$$

- Choose the value of τ^2 that makes the prior modes equal to $\psi(\omega, r)$.

Choice of r: Variability around effect sizes, Replicated studies

- Assume the prior on r is proportional to a Cauchy density truncated in the interval $(1, \infty)$ (denoted by $C_{1+}(r)$).

Choice of r: Variability around effect sizes, Replicated studies

- Assume the prior on r is proportional to a Cauchy density truncated in the interval $(1, \infty)$ (denoted by $C_{1+}(r)$).

- r can be estimated in several ways. Here, we propose the marginal maximum *a posteriori* (MMAP) estimate r^*_ω defined by

$$r^*_\omega = \arg\max_{r \geq 1} \left[\prod_{s=1}^{S} m_1(x_s \,|\, r, \tau^2_{\omega,r}) \right] \pi_N(r), \quad (6)$$

where $m_1(x_s \,|\, r, \tau^2_{\omega,r})$ represents the marginal density of the test statistic x_s, $s = 1, \ldots, S$ given ω and r.

Choice of r: Variability around effect sizes, Replicated studies

- Assume the prior on r is proportional to a Cauchy density truncated in the interval $(1, \infty)$ (denoted by $C_{1+}(r)$).

- r can be estimated in several ways. Here, we propose the marginal maximum *a posteriori* (MMAP) estimate r^*_ω defined by

$$r^*_\omega = \arg\max_{r \geq 1} \left[\prod_{s=1}^{S} m_1(x_s \,|\, r, \tau^2_{\omega,r}) \right] \pi_N(r), \tag{6}$$

  where $m_1(x_s \,|\, r, \tau^2_{\omega,r})$ represents the marginal density of the test statistic x_s, $s = 1, \ldots, S$ given ω and r.

- $r^*_\omega = 1$ for a single replication.

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence
- $(X, Y) \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$[Assumption not required for computing BFFs]

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence
- $(X, Y) \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$[Assumption not required for computing BFFs]
- Hypothesis: $H_0 : \omega($or $\rho) = 0$ against $H_1 : \omega($or $\rho) \geq 0$, ω: Standardized effect size

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence
- $(X,Y) \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$[Assumption not required for computing BFFs]
- Hypothesis: $H_0 : \omega$(or ρ) $= 0$ against $H_1 : \omega$(or ρ) ≥ 0, ω: Standardized effect size

Notations:

- Total number of replications = 20

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence
- $(X, Y) \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$[Assumption not required for computing BFFs]
- Hypothesis: $H_0 : \omega(\text{or } \rho) = 0$ against $H_1 : \omega(\text{or } \rho) \geq 0$, ω: Standardized effect size

Notations:

- Total number of replications = 20
- Denote r_i = Sample correlation coefficient and n_i = Sample size for the i-th replication, i = 1,2,...20.

Example: OSC data

Problem: Two variables:

- X: Conscientousness, Y: Persistence
- $(X, Y) \sim N_2(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$[Assumption not required for computing BFFs]
- Hypothesis: $H_0 : \omega(\text{or } \rho) = 0$ against $H_1 : \omega(\text{or } \rho) \geq 0$, ω: Standardized effect size

Notations:

- Total number of replications = 20
- Denote r_i = Sample correlation coefficient and n_i = Sample size for the i-th replication, i = 1,2,...20.
- $t_i = \frac{1}{2} \log\left(\frac{1+r_i}{1-r_i}\right)$= Fisher's transformation of r_i.

Example: Continued

- If ρ_i denotes the population correlation coefficient for the $i-th$ study, $t_i \sim N\left(\frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right), \frac{1}{n_i-3}\right)$.

Example: Continued

- If ρ_i denotes the population correlation coefficient for the $i-th$ study, $t_i \sim N\left(\frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right), \frac{1}{n_i-3}\right)$.

- Define $z_i = \sqrt{n_i - 3}t_i$. Therefore, $z_i \sim N(\lambda_i, 1)$, where $\lambda_i = \frac{\sqrt{n_i-3}}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right)$ is the non-centrality parameter.

Example: Continued

- If ρ_i denotes the population correlation coefficient for the $i-th$ study, $t_i \sim N\left(\frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right), \frac{1}{n_i-3}\right)$.
- Define $z_i = \sqrt{n_i - 3}t_i$. Therefore, $z_i \sim N(\lambda_i, 1)$, where $\lambda_i = \frac{\sqrt{n_i-3}}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right)$ is the non-centrality parameter.
- $\rho =$ Population correlation coefficient across all studies.

Example: Continued

- If ρ_i denotes the population correlation coefficient for the $i-th$ study, $t_i \sim N\left(\frac{1}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right), \frac{1}{n_i-3}\right)$.

- Define $z_i = \sqrt{n_i - 3}t_i$. Therefore, $z_i \sim N(\lambda_i, 1)$, where $\lambda_i = \frac{\sqrt{n_i-3}}{2}\log\left(\frac{1+\rho_i}{1-\rho_i}\right)$ is the non-centrality parameter.

- $\rho =$ Population correlation coefficient across all studies.

- $\omega = \frac{1}{2}\log\left(\frac{1+\rho}{1-\rho}\right) =$ Standardized effect size

Example(Continued): Prior and Choices of hyper-parameters

- Given ω, $\lambda_i \overset{iid}{\sim} \pi_{NM}(\mu \,|\, r, \tau^2_{r,\omega,i})$, $\tau^2_{r,\omega,i} = \frac{(n_i-3)\omega^2}{2r}$

Example(Continued): Prior and Choices of hyper-parameters

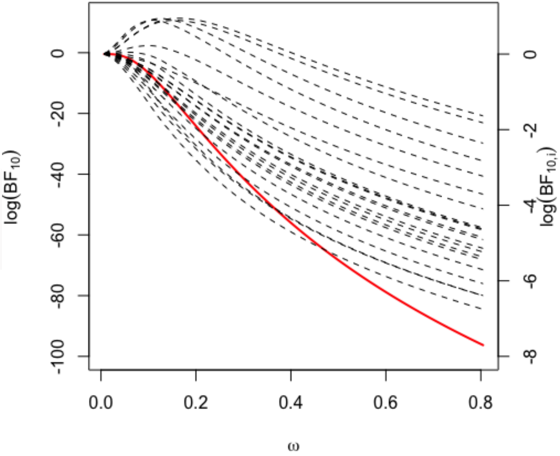
- Given ω, $\lambda_i \overset{iid}{\sim} \pi_{NM}(\mu \,|\, r, \tau^2_{r,\omega,i})$, $\tau^2_{r,\omega,i} = \frac{(n_i-3)\omega^2}{2r}$
- MAP estimate of $r = r^*_\omega$, assuming a half Cauchy prior on r.

Example(Continued): Prior and Choices of hyper-parameters

- Given ω, $\lambda_i \overset{iid}{\sim} \pi_{NM}(\mu \,|\, r, \tau^2_{r,\omega,i})$, $\tau^2_{r,\omega,i} = \frac{(n_i-3)\omega^2}{2r}$
- MAP estimate of $r = r^*_\omega$, assuming a half Cauchy prior on r.
- The Bayes factors based on the 20 replications of the experiment, given r and ω, can be expressed as the product of Bayes factors from the individual experiments. Applying Theorem 2.5,

$$BF_{10}(z \,|\, \omega, r) = \prod_{i=1}^{20} BF_{10}(z_i | \tau^2_{\omega,r,i}, r). \tag{7}$$

Combined BFF and Individual BFF of replication studies



Observations

- Very strong evidence in favour of the null hypothesis

Observations

- Very strong evidence in favour of the null hypothesis
- $\log(BF_{10})$ centered on effect sizes greater than $\rho = 0.20$ were less than -32.

Observations

- Very strong evidence in favour of the null hypothesis
- $\log(BF_{10})$ centered on effect sizes greater than $\rho = 0.20$ were less than -32.
- For these data, $r^*_\omega = 1$ for $\omega \in (0, 0.082) \cup (0.150, \infty)$ and did not exceed 1.172 in the interval $(0.082, 0.150)$.

Observations

- Very strong evidence in favour of the null hypothesis
- $\log(BF_{10})$ centered on effect sizes greater than $\rho = 0.20$ were less than -32.
- For these data, $r^*_\omega = 1$ for $\omega \in (0, 0.082) \cup (0.150, \infty)$ and did not exceed 1.172 in the interval $(0.082, 0.150)$.
- This is due to the fact that the null is favoured in this study.

Standard choice of r

When there is no prior information about the dispersion of the non-centrality parameter across several replications, choose $r = 1$.

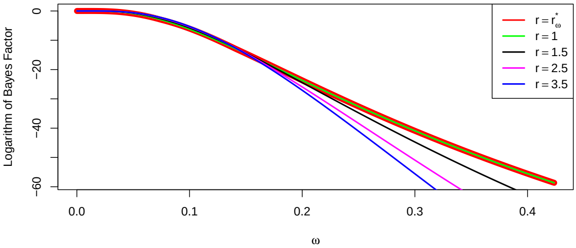


Figure: BFF for various values of r

Comparison with other standard methods

Competing method(Ly, Verhagen, Wagenmakers,2016)
Assuming a Bivariate normal model,for the $i-th$ replication:

- $\rho_i \sim$ Stretched-beta$(1/\kappa, 1/\kappa)$.
- $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \propto \frac{1}{\sigma_x \sigma_y}$

Comparison with other standard methods

Competing method(Ly, Verhagen, Wagenmakers,2016)
Assuming a Bivariate normal model,for the $i-th$ replication:

- $\rho_i \sim$ Stretched-beta$(1/\kappa, 1/\kappa)$.
- $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \propto \frac{1}{\sigma_x \sigma_y}$

Mode of Comparison:

- Obtain the maximum Bayes Factor(max BF_{10}) for each study by maximizing with respect to κ

Comparison with other standard methods

Competing method(Ly, Verhagen, Wagenmakers,2016)
Assuming a Bivariate normal model,for the $i-th$ replication:

- $\rho_i \sim$ Stretched-beta$(1/\kappa, 1/\kappa)$.
- $\pi(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2) \propto \frac{1}{\sigma_x \sigma_y}$

Mode of Comparison:

- Obtain the maximum Bayes Factor(max BF_{10}) for each study by maximizing with respect to κ
- We obtain the maximum BF using Bayes Factor function for each study(using $r = 1$).

Comparison of maximum BFs

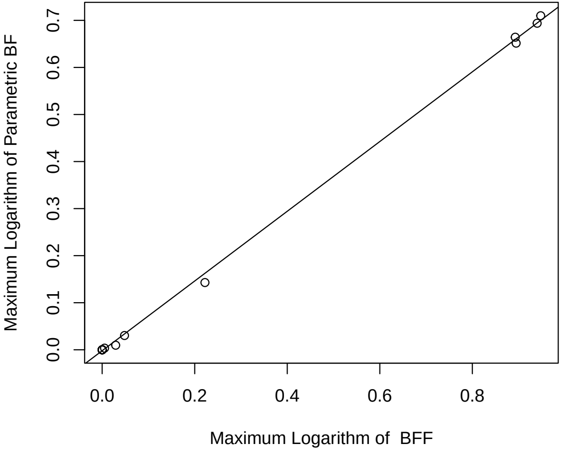
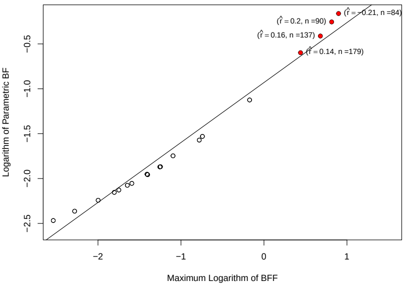
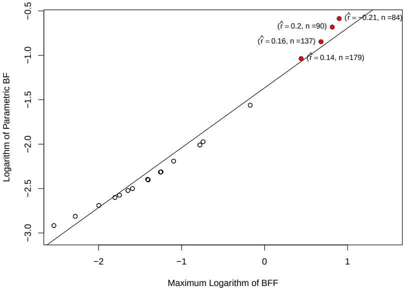


Figure: Unrestricted ω and κ

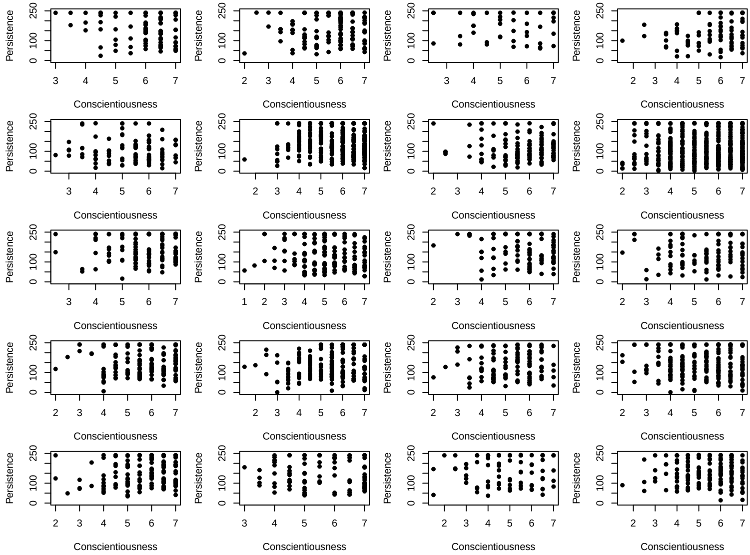
Comparison of maximum BFs





Some drawbacks of the fully parametric model

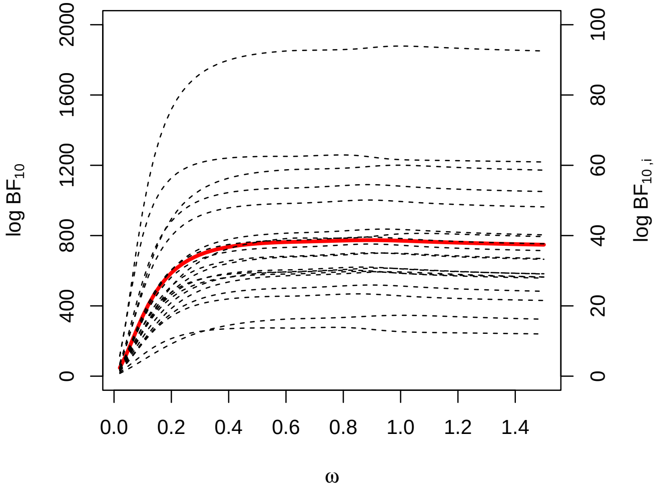
Assumes normality whereas the underlying distribution of the data is not normal.



Example:Stroop test

- A test for the difference of means between two populations
- A frequentist t test is done
- Impose a $NM^{+}(\tau^2, r)$ prior on the non-centrality parameter of the t-test statistic under the alternative
- Standardized effect size(ω) $= \frac{(\mu_1 - \mu_2)}{\sigma}$
- $\tau^2 = \frac{n_1 n_2 \omega^2}{2r(n_1+n_2)}, r = r^*_\omega$
- Replications $= 36$

Combined and Individual BFFs



Observations

- Overwhelming support in favour of the alternative

Observations

- Overwhelming support in favour of the alternative
- $\log(\text{BFF}) = 774$ at $\omega = 0.9$

Observations

- Overwhelming support in favour of the alternative
- log(BFF) $= 774$ at $\omega = 0.9$
- $r^*_\omega = 12$.

BFF for varying r

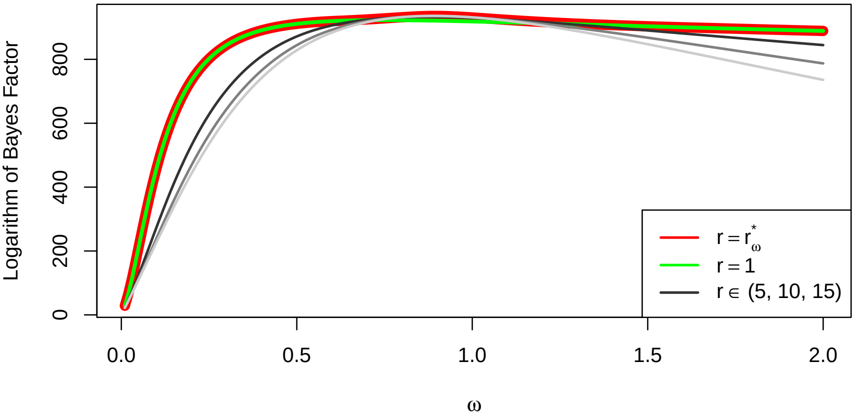
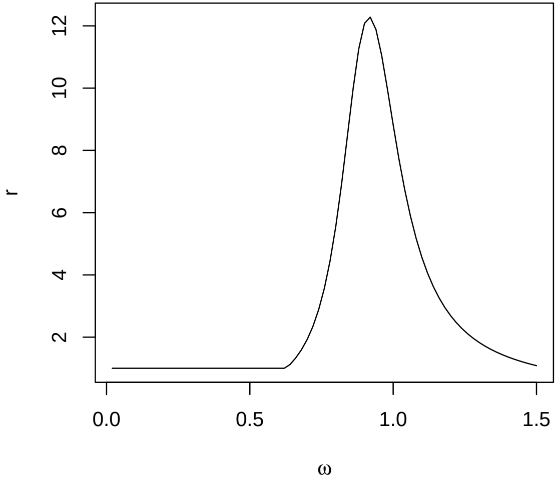


Figure: Bayes factor functions for various values of r

Choice of r



Conclusion

- Overcomes computational complexities of Bayes factors by defining Bayes factors from classical test statistics and using standardized effect sizes to define alternative hypotheses

Conclusion

- Overcomes computational complexities of Bayes factors by defining Bayes factors from classical test statistics and using standardized effect sizes to define alternative hypotheses
- Reflects effect sizes by expressing Bayes factors as functions of effect sizes

Conclusion

- Overcomes computational complexities of Bayes factors by defining Bayes factors from classical test statistics and using standardized effect sizes to define alternative hypotheses
- Reflects effect sizes by expressing Bayes factors as functions of effect sizes
- Accounts for dispersion of the effect sizes and draws sensible conclusion under replicated design

Conclusion

- Overcomes computational complexities of Bayes factors by defining Bayes factors from classical test statistics and using standardized effect sizes to define alternative hypotheses
- Reflects effect sizes by expressing Bayes factors as functions of effect sizes
- Accounts for dispersion of the effect sizes and draws sensible conclusion under replicated design
- Enhances interpretation of Bayes factors by centering the modes of the alternative prior density on values determined by standardized effect size and hence overcoming the subjectivity of the priors

References

- *Valen E. Johnson and Sandipan Pramanik and Rachael Shudde*,Bayes factor functions for reporting outcomes of hypothesis tests,*PNAS 2023*
- *Valen E. Johnson*, Bayes Factors Based on Test Statistics, *JRSS B, 2005*
- *Alexander Ly , Josine Verhagen, Eric-Jan Wagenmakers*,Harold Jeffreys's default Bayes factor hypothesis tests: Explanation, extension, and application in psychology,*Journal of Mathematical Psychology, 2016*

Thank You